Supplementary Appendices: Landmark Estimation of Survival and Treatment Effect in a Randomized Clinical Trial

Layla Parast[1], Lu Tian[2] and Tianxi Cai[3]

[1]RAND Corporation, Santa Monica, CA 90401

[2] Stanford University, Department of Health, Research and Policy, Stanford, CA 94305

[3]Harvard University, Department of Biostatistics, Boston, MA 02115

Supplementary Appendices

Recall that our proposed estimate $\widehat{S}_{\text{LM}}(t) = \widehat{S}(t \mid t_0)\widehat{S}(t_0)$. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^*$ be the maximizers of the log partial likelihood functions corresponding to the proportional hazards working models (2.3) and (2.4), respectively. In addition, let $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_0^*$ denote the limits of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^*$, respectively. Let $\widehat{S}(\boldsymbol{\beta}^*, t_0) = n^{-1}\sum_{i=1}^n \widehat{S}_{\mathbf{Z}_i'\boldsymbol{\beta}^*}(\boldsymbol{\beta}^*, t_0)$, $\widehat{S}(\boldsymbol{\beta}, t \mid t_0) = n_{t_0}^{-1}\sum_{i=1}^{n_{t_0}} \widehat{S}_{\mathbf{W}_i'\boldsymbol{\beta}}(\boldsymbol{\beta}, t \mid t_0)$, and $\widehat{S}_{\text{LM}}(\boldsymbol{\beta}^*, \boldsymbol{\beta}, t) = \widehat{S}(\boldsymbol{\beta}^*, t_0)\widehat{S}(\boldsymbol{\beta}, t \mid t_0)$, where $\widehat{S}_u(\boldsymbol{\beta}^*, t_0) = e^{-\widehat{\Lambda}_u(\boldsymbol{\beta}^*, t_0)}$ $\widehat{S}_v(\boldsymbol{\beta}, t \mid t_0) = e^{-\widehat{\Lambda}_v(\boldsymbol{\beta}, t|t_0)}$, $\widehat{\Lambda}_u(\boldsymbol{\beta}^*, t_0)$ and $\widehat{\Lambda}_v(\boldsymbol{\beta}, t \mid t_0)$ are obtained by replacing $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\beta}}$ in $\widehat{\Lambda}_u(t_0)$ and $\widehat{\Lambda}_v(t \mid t_0)$ by $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}$, respectively. Then $\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0) = \widehat{S}(t_0)$, $\widehat{S}(\widehat{\boldsymbol{\beta}}, t \mid t_0) = \widehat{S}(t \mid t_0)$, and $\widehat{S}_{\text{LM}}(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\beta}}, t) = \widehat{S}_{\text{LM}}(t)$. Furthermore, with a slight abuse of notation, we define $S_u(\boldsymbol{\beta}^*, t_0) = P(T_{\mathbb{L}i} > t_0 | \mathbf{Z}_i'\boldsymbol{\beta}^* = u) = e^{-\Lambda_u(\boldsymbol{\beta}^*, t_0)}$ and $S_v(\boldsymbol{\beta}, t \mid t_0) = P(T_{\mathbb{L}i} > t | \mathbf{W}_i'\boldsymbol{\beta} = v, T_{\mathbb{L}i} > t_0) = e^{-\Lambda_v(\boldsymbol{\beta}, t|t_0)}$ Hence $S_u(t_0) = S_u(\boldsymbol{\beta}_0^*, t_0)$ and $S_v(t \mid t_0) = S_v(\boldsymbol{\beta}_0, t \mid t_0)$. Let $\widehat{\mathcal{W}}_{\text{LM}}(t, t_0) \equiv \sqrt{n}\{\widehat{S}_{\text{LM}}(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\beta}}, t) - S(t)\}$, which can be decomposed as

$$\sqrt{n}\{\widehat{S}_{\text{LM}}(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\beta}}, t) - \widehat{S}_{\text{LM}}(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}_0, t)\} + \sqrt{n}\{\widehat{S}_{\text{LM}}(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}_0, t) - S(t)\}$$

and let $\widehat{\mathcal{W}}^{\mathbf{Z}}(t_0) = \sqrt{n}\{\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0) - S(t_0)\}$ and $\widehat{\mathcal{W}}^{\mathbf{W}}(t \mid t_0) = \sqrt{n_{t_0}}\{\widehat{S}(\widehat{\boldsymbol{\beta}}, t|t_0) - S(t|t_0)\}$.

Assumption A.1 Censoring, C, is independent of $\{T_{\mathbb{L}}, \mathbf{Z}\}$ given treatment assignment, $\mathcal{G}$.

Assumption A.2 Treatment assignment, $\mathcal{G}$, is independent of $\mathbf{Z}$.

Regularity Conditions (C.1) Throughout, we assume several regularity conditions: (1) the joint density of $\{T_{\mathbb{L}}, \mathbf{T}_{\mathbb{S}}, C\}$ is continuously differentiable with density function bounded away from 0 on $[0, t]$, (2) $\mathbf{Z}^{\mathsf{T}}\boldsymbol{\beta}_0^*$ and $\mathbf{W}^{\mathsf{T}}\boldsymbol{\beta}_0$ have continuously differentiable densities and $\mathbf{Z}$ is bounded, (3) $h = O(n^{-v})$ with $1/4 < v < 1/2$, (4) $K(x)$ is a symmetric smooth kernel function with continuous second derivative on its support and $\int \dot{K}(x)^2 dx < \infty$, where $\dot{K}(x) = dK(x)/dx$.

Supplementary Appendix A: Variability of $\widehat{\boldsymbol\beta}$ and $\widehat{\boldsymbol\beta}^*$

In Theorem S.1 we show that the variability of $\widehat{\boldsymbol\beta}^*$ and $\widehat{\boldsymbol\beta}$ is negligible and can be ignored in making inferences on $\widehat{S}(\widehat{\boldsymbol\beta}^*, t_0)$, $\widehat{S}(\widehat{\boldsymbol\beta}, t|t_0)$ and $\widehat{S}_{\rm LM}(\boldsymbol\beta_0^*, \boldsymbol\beta_0, t)$.

Theorem S.1. *Assume A.1, A.2, and Regularity Conditions C.1 hold. Then*

$$\sqrt{n}\{\widehat{S}_{LM}(\widehat{\boldsymbol\beta}^*, \widehat{\boldsymbol\beta}, t) - \widehat{S}_{LM}(\boldsymbol\beta_0^*, \boldsymbol\beta_0, t)\} = o_p(1)$$

Proof of Theorem S.1. We focus on $\widehat{S}(\widehat{\boldsymbol\beta}^*, t_0)$ only as properties of $\widehat{S}(\widehat{\boldsymbol\beta}, t|t_0)$ follow similarly. We will show that the variability of $\widehat{\boldsymbol\beta}^*$ is negligible in the sense that $\sqrt{n}\{\widehat{S}(\widehat{\boldsymbol\beta}^*, t_0) - \widehat{S}(\boldsymbol\beta_0^*, t_0)\} = o_p(1)$. Let $\widehat{F}(\boldsymbol\beta, u) = n^{-1}\sum_{i=1}^n I(\boldsymbol\beta'\mathbf{Z}_i \le u)$ and $F(\boldsymbol\beta, u) = P(\boldsymbol\beta'\mathbf{Z} \le u)$. We estimate $S(t_0) = P(T_{\mathbb{L}} > t_0)$ by $\widehat{S}(\widehat{\boldsymbol\beta}^*, t_0) = \int \widehat{S}_u(\widehat{\boldsymbol\beta}^*, t_0)\widehat{F}(\widehat{\boldsymbol\beta}^*, du)$. For $\|\boldsymbol\beta_2 - \boldsymbol\beta_1\| \le \epsilon = o(1)$, we then have $\sqrt{n}\{\widehat{S}(\boldsymbol\beta_2, t_0) - \widehat{S}(\boldsymbol\beta_1, t_0)\} = \sqrt{n}[I_1+I_2+I_3+I_4]$, where $I_1 = \int\{\widehat{S}_u(\boldsymbol\beta_2, t_0) - \widehat{S}_u(\boldsymbol\beta_1, t_0) - S_u(\boldsymbol\beta_2, t_0) + S_u(\boldsymbol\beta_1, t_0)\}\widehat{F}(\boldsymbol\beta_2, du)$, $I_2 = \int\{S_u(\boldsymbol\beta_2, t_0) - S_u(\boldsymbol\beta_1, t_0)\}\{\widehat{F}(\boldsymbol\beta_2, du) - F(\boldsymbol\beta_2, du)\}$, $I_3 = \int \widehat{S}_u(\boldsymbol\beta_1, t_0)\{\widehat{F}(\boldsymbol\beta_2, du) - \widehat{F}(\boldsymbol\beta_1, du) - F(\boldsymbol\beta_2, du) + F(\boldsymbol\beta_1, du)\}$, and $I_4 = \int\{\widehat{S}_u(\boldsymbol\beta_1, t_0) - S_u(\boldsymbol\beta_1, t_0)\}\{F(\boldsymbol\beta_2, du) - F(\boldsymbol\beta_1, du)\}$. Here we used the fact that

$$\int S_u(\boldsymbol\beta_2, t_0)F(\boldsymbol\beta_2, du) = \int S_u(\boldsymbol\beta_1, t_0)F(\boldsymbol\beta_1, du) = S(t_0).$$

We next bound I_1, I_2, I_3 and I_4. To bound I_1, we will show that

$$\sup_{u,\boldsymbol\beta}\left|\left\{\frac{\partial\widehat{S}_u(\boldsymbol\beta, t_0)}{\partial\boldsymbol\beta} - \frac{\partial S_u(\boldsymbol\beta, t_0)}{\partial\boldsymbol\beta}\right\}\right| = o_p(1). \tag{A·1}$$

First, $\sup_{u,\boldsymbol\beta}|\partial\widehat{S}_u(\boldsymbol\beta, t_0)/\partial\boldsymbol\beta - \partial S_u(\boldsymbol\beta, t_0)/\partial\boldsymbol\beta| \le \tilde{C}_1\sup_{u,\boldsymbol\beta}|\partial\widehat{\Lambda}_u(\boldsymbol\beta, t_0)/\partial\boldsymbol\beta - \partial\Lambda_u(\boldsymbol\beta, t_0)/\partial\boldsymbol\beta\}|$, where $\tilde{C}_1$ is a positive constant and

$$\frac{\partial\widehat{\Lambda}_u(\boldsymbol\beta, t_0)}{\partial\boldsymbol\beta} = \int_0^{t_0}\frac{d\left\{n^{-1}\sum_{i=1}^n N_i(t)h^{-1}\dot{K}_h(\boldsymbol\beta'\mathbf{Z}_i - u)\mathbf{Z}_i\right\}}{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol\beta'\mathbf{Z}_i - u)}$$

$$-\int_0^{t_0} \frac{d\left\{n^{-1}\sum_{i=1}^n N_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)\right\}\left\{n^{-1}\sum_{i=1}^n Y_i(t)h^{-1}\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)\mathbf{Z}_i\right\}}{\left\{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)\right\}^2}.$$

Note that

$$\begin{aligned}
n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i-u) - \mathrm{E}\{Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)\} &= \int I(x\geq t)K_h(s-u)d\{\widehat{F}_{\boldsymbol{\beta}}(x,s)-F_{\boldsymbol{\beta}}(x,s)\}\\
=n^{-1/2}\int I(x\geq t)K_h(s-u)dG_{F_n}(x,s;\boldsymbol{\beta}) + O\{n^{-1}h^{-1}\log(n)^2\} &= o\{(n^{-1/2}+(nh)^{-1})n^{\epsilon}\},
\end{aligned}$$

for any $\epsilon>0$, where $F_{\boldsymbol{\beta}}(x,s)=\mathrm{P}(X_i\leq x,\boldsymbol{\beta}'\mathbf{Z}_i\leq s)$, $\widehat{F}_{\boldsymbol{\beta}}(x,s;\boldsymbol{\beta})=n^{-1}\sum_{i=1}^n I(X_i\leq x,\boldsymbol{\beta}'\mathbf{Z}_i\leq s)$ and $G_{F_n}(x,s;\boldsymbol{\beta})$ is a Gaussian process such that

$$\sup_{x,s,\boldsymbol{\beta}}\|\sqrt{n}\{\widehat{F}_{\boldsymbol{\beta}}(x,s)-F_{\boldsymbol{\beta}}(x,s)\}-G_{F_n}(x,s;\boldsymbol{\beta})\| = O(n^{-1/2}\log(n)^2)\quad \text{almost surely.}$$

The existence of such a Gaussian process, which is a time-transformed Brownian bridge, is ensured by the strong approximation result of Tusnády (1977). In the last step above, we used the fact that $\sup_{u,\boldsymbol{\beta}}\|\int I(x\geq t)K_h(s-u)dG_{F_n}(x,s;\boldsymbol{\beta})\| = o(n^{\epsilon})$ for any $\epsilon>0$ (Bickel & Rosenblatt, 1973). Therefore, we have

$$\sup_{\boldsymbol{\beta},u}\left| n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)-\mathrm{E}\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i=u\}f_{\boldsymbol{\beta}'Z}(v)\right| = O\{(n^{-1/2}+(nh)^{-1})n^{\epsilon}+h^2\}.$$

for any $\epsilon>0$. Similarly, for any $\epsilon>0$,

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n Y_i(t)h^{-1}\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)Z_{1i} - \mathrm{E}\{Y_i(t)h^{-1}\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i-u)Z_{1i}\}\\
=&n^{-1/2}\int I(x\geq t)zh^{-1}\dot{K}_h(s-u)dG_{H_n}(x,s,z;\boldsymbol{\beta}) + O\{h^{-1}n^{-2/3}\log(n)^{\tilde{d}}\} = o(n^{\epsilon-1/2}h^{-1}),
\end{aligned}$$

where $\tilde{d}>0$, Z_{1i} is the first component of vector $\mathbf{Z}_i$, $H_{\boldsymbol{\beta}}(x,s,z)=\mathrm{P}(X_i\leq x,\boldsymbol{\beta}'\mathbf{Z}\leq s,Z_{1i}\leq z)$

$\widehat{H}_{\boldsymbol\beta}(x,s,z)=n^{-1}\sum_{i=1}^n I(X_i\le x,\boldsymbol\beta'\mathbf{Z}_i\le s,Z_{1i}\le z)$ and $G_{H_n}(x,s;\boldsymbol\beta)$ is a Gaussian process such that

$$\sup_{x,s,z,\boldsymbol\beta}\|\sqrt{n}\{\widehat{H}_{\boldsymbol\beta}(x,s,z)-H_{\boldsymbol\beta}(x,s,z)\}-G_{H_n}(x,s,z;\boldsymbol\beta)\|=O(n^{-1/6}\log(n)^{\tilde d})\quad\text{almost surely.}$$

The existence of the Gaussian process is ensured by the results of Massart (1989). Furthermore, by the standard Taylor series expansion, we have

$$\sup_{\boldsymbol\beta,u}\left|\,n^{-1}\sum_{i=1}^n Y_i(t)h^{-1}\dot K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\mathbf{Z}_i-\frac{\partial\mathrm{E}\{Y_i(t)|\boldsymbol\beta'\mathbf{Z}_i=u\}}{\partial\boldsymbol\beta}f_{\boldsymbol\beta'Z}(v)\,\right|=O(n^{\epsilon-1/2}h^{-1}+h).$$

for any $\epsilon>0$. Similarly, we can show that

$$\sup_{\boldsymbol\beta,u}\left|\,n^{-1}\sum_{i=1}^n N_i(t)h^{-1}\dot K_h(\boldsymbol\beta'\mathbf{Z}_i-u)Z_i-\frac{\partial\mathrm{E}\{N_i(t)|\boldsymbol\beta'\mathbf{Z}_i=u\}}{\partial\boldsymbol\beta}f_{\boldsymbol\beta'Z}(u)\,\right|=O(n^{\epsilon-1/2}h^{-1}+h)$$

and $\sup_{\boldsymbol\beta,u}|n^{-1}\sum_{i=1}^n N_i(t)K_h(\boldsymbol\beta'\mathbf{Z}_i-u)-\mathrm{E}\{N_i(t)|\boldsymbol\beta'\mathbf{Z}_i=u\}f_{\boldsymbol\beta'Z}(u)|=O\{(n^{-1/2}+(nh)^{-1})n^{\epsilon}+h^2\}$. Furthermore, using integration by parts we have

$$\begin{aligned}\frac{\partial\widehat\Lambda_u(\boldsymbol\beta,t_0)}{\partial\boldsymbol\beta}&=\frac{\sum_{i=1}^n N_i(t_0)h^{-1}\dot K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\mathbf{Z}_i}{\sum_{i=1}^n Y_i(t_0)K_h(\boldsymbol\beta'\mathbf{Z}_i-u)}\\&-\int_0^{t_0}n^{-1}\sum_{i=1}^n N_i(t)h^{-1}\dot K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\mathbf{Z}_i d\left\{\frac{1}{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol\beta'\mathbf{Z}_i-u)}\right\}\\&-\int_0^{t_0}\frac{\left\{n^{-1}\sum_{i=1}^n Y_i(t)h^{-1}\dot K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\mathbf{Z}_i\right\}d\left\{n^{-1}\sum_{i=1}^n N_i(t)K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\right\}}{\left\{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol\beta'\mathbf{Z}_i-u)\right\}^2}.\end{aligned}$$

Thus $\partial\widehat{\Lambda}_u(\boldsymbol{\beta},t_0)/\partial\boldsymbol{\beta} - \partial\Lambda_u(\boldsymbol{\beta},t_0)/\partial\boldsymbol{\beta}$ can be bounded by

$$
\begin{aligned}
& \left| \frac{\sum_{i=1}^n N_i(t_0)h^{-1}\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)\mathbf{Z}_i}{\sum_{i=1}^n Y_i(t_0)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)} - \frac{\partial E\{N_i(t_0)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}/\partial\boldsymbol{\beta}}{E\{Y_i(t_0)|\boldsymbol{\beta}'\mathbf{Z}_i = u)\}} \right| \\
& + \int_0^{t_0} \left| \frac{1}{nh}\sum_{i=1}^n N_i(t)\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)\mathbf{Z}_i - \frac{\partial E\{N_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}}{\partial\boldsymbol{\beta}} f_{\boldsymbol{\beta}'Z}(u) \right| d\left\{ \frac{1}{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)} \right\} \\
& + \left| \int_0^{t_0} \frac{\partial E\{N_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}}{\partial\boldsymbol{\beta}} f_{\boldsymbol{\beta}'Z}(u) d\left\{ \frac{1}{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)} - \frac{1}{E\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}f_{\boldsymbol{\beta}'V}(u)} \right\} \right| \\
& + \int_0^{t_0} \left| \frac{n^{-1}\sum_{i=1}^n Y_i(t)h^{-1}\dot{K}_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)\mathbf{Z}_i}{\left\{n^{-1}\sum_{i=1}^n Y_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u)\right\}^2} - \frac{\partial E\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}/\partial\boldsymbol{\beta}}{E\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}^2 f_{\boldsymbol{\beta}'Z}(u)} \right| d\left\{ n^{-1}\sum_{i=1}^n N_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u) \right\} \\
& + \left| \int_0^{t_0} \frac{\partial E\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i - u)/\partial\boldsymbol{\beta}}{E\{Y_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}^2 f_{\boldsymbol{\beta}'Z}(u)} d\left\{ n^{-1}\sum_{i=1}^n N_i(t)K_h(\boldsymbol{\beta}'\mathbf{Z}_i - u) - E\{N_i(t)|\boldsymbol{\beta}'\mathbf{Z}_i = u\}f_{\boldsymbol{\beta}'V}(u) \right\} \right|
\end{aligned}
$$

$=O(n^{\epsilon-1/2}h^{-1} + h) = o_p(1)$ for any $\epsilon > 0$, where $h = O(n^{-\delta}), \delta \in (0, 1/2)$.

Therefore, (A·1) holds. It then follows that

$$
\begin{aligned}
I_1 &= \int \left\{ \frac{\partial\widehat{S}_u(\boldsymbol{\beta}_2, t_0)}{\partial\beta} - \frac{\partial S_u(\boldsymbol{\beta}_2, t_0)}{\partial\beta} \right\} \widehat{F}(\boldsymbol{\beta}_2, du) O_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|) \\
&= \int \left\{ \frac{\partial\widehat{S}_u(\boldsymbol{\beta}_2, t_0)}{\partial\beta} - \frac{\partial S_u(\boldsymbol{\beta}_2, t_0)}{\partial\beta} \right\} \{F(\boldsymbol{\beta}_2, du) + O_p(n^{\frac{1}{2}})\} O_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|) = o_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|).
\end{aligned}
$$

Next, we note that $I_2 = \int\{S_u(\boldsymbol{\beta}_2, t_0) - S_u(\boldsymbol{\beta}_1, t_0)\}\{\widehat{F}(\boldsymbol{\beta}_2, du) - F(\boldsymbol{\beta}_2, du)\} = O_p(n^{-1/2}|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|)$, and

$$
\begin{aligned}
I_3 &= \int \widehat{S}_u(\boldsymbol{\beta}_1, t_0)\{\widehat{F}(\boldsymbol{\beta}_2, du) - \widehat{F}(\boldsymbol{\beta}_1, du) - F(\boldsymbol{\beta}_2, du) + F(\boldsymbol{\beta}_1, du)\} \\
&\leq \sup_{\boldsymbol{\beta}} |\widehat{S}(\boldsymbol{\beta}, t_0)| \int |\widehat{F}(\boldsymbol{\beta}_2, u) - \widehat{F}(\boldsymbol{\beta}_1, u) - F(\boldsymbol{\beta}_2, u) + F(\boldsymbol{\beta}_1, u)| du = O_p(n^{-1/2}|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|^{1/2}) \\
I_4 &= \sup_{\boldsymbol{\beta}, u} |\widehat{S}_u(\boldsymbol{\beta}_1, t_0) - S_u(\boldsymbol{\beta}_1, t_0)| \times O_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|) = o_p(\{(nh)^{-1/2}\log(n) + h^2\}|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|).
\end{aligned}
$$

where we used the fact that $\widehat{S}(\boldsymbol{\beta}_1, t_0) = O_p(1)$ and $\sup_{|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2| \leq \delta, u} n^{1/2}|\widehat{F}(\boldsymbol{\beta}_2, u) - \widehat{F}(\boldsymbol{\beta}_1, u) - F(\boldsymbol{\beta}_2, u) + F(\boldsymbol{\beta}_1, u)| = O_p(\delta^{1/2})$. Therefore, $n^{1/2}(I_1 + I_2 + I_3 + I_4)$ is bounded by $o_p(n^{1/2}|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1| + O_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|^{1/2})$

for $h = O_p(n^{-\delta}), \delta \in (0, 1/2)$. This implies that

$$\sqrt{n}\{\widehat{S}(\boldsymbol{\beta}_2, t_0) - \widehat{S}(\boldsymbol{\beta}_1, t_0)\} = o_p(n^{1/2}|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1| + O_p(|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1|^{1/2}).$$

Since this result is uniform for $(\boldsymbol{\beta}_2, \boldsymbol{\beta}_1)$, we can let $(\boldsymbol{\beta}_2, \boldsymbol{\beta}_1) = (\widehat{\boldsymbol{\beta}}^*, \boldsymbol{\beta}_0^*)$ and obtain

$$\sqrt{n}\{\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0) - \widehat{S}(\boldsymbol{\beta}_0^*, t_0)\} = o_p(1),$$

where we used the fact that $|\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0^*| = O_p(n^{-1/2})$. Therefore, the variability of $\widehat{\boldsymbol{\beta}}^*$ can be ignored in making inferences on $\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0)$ and $\widehat{\mathcal{W}}^{\mathbf{z}}(t_0)$ is asymptotically equivalent to $\widetilde{\mathcal{W}}^{\mathbf{z}}(t_0) \equiv \sqrt{n}\{\widehat{S}(\boldsymbol{\beta}_0^*, t_0) - S(t_0)\}$. Note that the derivations above only requires $h = O(n^{-\delta})$ with $\delta \in (0, 1/2)$. Intuitively, the reason $\widehat{\boldsymbol{\beta}}^*$ does not contribute any additional noise to $\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0)$ is due to the fact that the limiting quantity $S(\boldsymbol{\beta}, t_0)$ is in fact free of $\boldsymbol{\beta}$. Our derivations essentially establish the stochastic equicontinuity property of $n^{\frac{1}{2}}\{\widehat{S}(\boldsymbol{\beta}, t_0) - S(\boldsymbol{\beta}, t_0)\}$. As a result,

$$\begin{aligned} n^{\frac{1}{2}}\{\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0) - \widehat{S}(\boldsymbol{\beta}_0^*, t_0)\} &= n^{\frac{1}{2}}\{S(\widehat{\boldsymbol{\beta}}^*, t_0) - S(\boldsymbol{\beta}_0^*, t_0)\} + o_p(1) \\ &= n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}_0^*)^{\mathsf{T}} \frac{\partial S(\boldsymbol{\beta}, t_0)}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0^*} + o_p(1). \end{aligned}$$

Since $S(\boldsymbol{\beta}, t_0)$ is free of $\boldsymbol{\beta}$, $\partial S(\boldsymbol{\beta}, t_0)/\partial \boldsymbol{\beta} = 0$ and hence $n^{\frac{1}{2}}\{\widehat{S}(\widehat{\boldsymbol{\beta}}^*, t_0) - \widehat{S}(\boldsymbol{\beta}_0^*, t_0)\} = o_p(1)$. Using the same arguments, it can be shown $\widehat{\mathcal{W}}^{\mathbf{w}}(t \mid t_0) \approx \widetilde{\mathcal{W}}^{\mathbf{w}}(t \mid t_0)$, where $\widetilde{\mathcal{W}}^{\mathbf{w}}(t \mid t_0) = \sqrt{n}\{\widehat{S}(\boldsymbol{\beta}_0, t|t_0) - S(t \mid t_0)\}$. Using a Taylor series expansion, one can show that

$$\widehat{\mathcal{W}}_{\text{LM}}(t, t_0) \approx S(t_0)\widetilde{\mathcal{W}}^{\mathbf{w}}(t|t_0) + S(t|t_0)\widetilde{\mathcal{W}}^{\mathbf{z}}(t_0). \tag{A·2}$$

It follows that $\widehat{\mathcal{W}}_{\text{LM}}(t, t_0)$ is asymptotically equivalent to $\widetilde{\mathcal{W}}_{\text{LM}}(t, t_0) \equiv \sqrt{n}\{\widehat{S}_{\text{LM}}(\boldsymbol{\beta}_0^*, \boldsymbol{\beta}_0, t) - S(t)\}$. Hence, we focus on the asymptotic expansions of $\widetilde{\mathcal{W}}^{\mathbf{z}}(t_0)$, $\widetilde{\mathcal{W}}(t \mid t_0)$, and $\widetilde{\mathcal{W}}_{\text{LM}}(t, t_0)$ with $\boldsymbol{\beta}_0^*$ and $\boldsymbol{\beta}_0$ as given. $\square$

Supplementary Appendix B: Relationship to Augmentation Approach

We now examine the relationship between our proposed estimation procedure and an augmentation approach. In Supplementary Appendix B.1 we show that our proposed estimate of survival can be expressed in the form of the Kaplan Meier estimate of survival plus an augmentation term composed of a martingale residual for censoring. In Supplementary Appendix B.2 we investigate the choice of basis function, $\mathbf{H}(\mathbf{Z}_i)$, when estimating $\widehat{\Delta}_{\text{AUG}}(t)$ i.e. for augmentation using covariates which are independent of treatment assignment. We show that an explicit form of the optimal basis can be obtained and estimated.

Supplementary Appendix B.1: Censoring Augmentation Term

To compare $\widehat{S}_{\text{LM}}(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\beta}}, t_0)$ to the KM estimate, we may approximate $\sqrt{n}\{\widehat{S}_{\text{KM}}(t) - S(t)\}$ by

$$
\begin{aligned}
&n^{-1/2} \sum_{i=1}^{n} \Bigg\{ I(X_{\mathbb{L}i} > t_0) S_C(t_0)^{-1} \left[S(t|t_0) \int_{t_0}^{t} \frac{dM_{C_i}(u|t_0)}{\pi(u|t_0)} + \frac{I(X_i > t)}{S_C(t|t_0)} - S(t|t_0) \right] \\
&+ S(t|t_0) \left[S(t_0) \int_{0}^{t_0} \frac{dM_{C_i}(u)}{\pi(u)} + \frac{I(T_i > t, C_i > t_0)}{S_C(t_0)} - S(t_0) \right] \Bigg\}
\end{aligned}
$$

This, together with (I.5) in Appendix I, implies that $\widehat{S}_{\text{LM}}(\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\beta}}, t) = \widehat{S}_{\text{KM}}(t) + n_{t_0}^{-1} \sum_{i \in \Omega_{t_0}} \gamma_{i1}(t, t_0) + n^{-1} \sum_{i=1}^{n} \gamma_{j2}(t, t_0) + o_p(n^{-\frac{1}{2}})$, where

$$
\begin{aligned}
\gamma_{i1}(t, t_0) &= S(t_0) \int_{t_0}^{t} \left[S_{U_i}(t \mid t_0) \frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)} - S(t|t_0) \frac{dM_{C_i}(u|t_0)}{\pi(u|t_0)} \right] \\
\gamma_{i2}(t, t_0) &= S(t|t_0) \int_{0}^{t_0} \left[S_{U_i^*}(t_0) \frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)} - S(t_0) \frac{dM_{C_i}(u)}{\pi(u)} \right].
\end{aligned}
$$

Thus, our complete proposed estimator can be expressed as the KM estimator plus an augmentation term composed of two quantities, one involving censoring occurring before t_0 and the other involving censoring occurring between t_0 and t.

Supplementary Appendix B.2: Covariate Augmentation Term for Estimating Treatment Difference

We next derive the optimal basis for augmenting $\widehat{\Delta}_{\text{LM}}(t)$ leveraging information from the randomized treatment assignment. Since only $\mathbf{Z}$, which is measured at baseline, is independent of treatment assignment, $\mathcal{G}$, (Assumption A.2) we only consider the class of functions $\{\mathbf{H}(\mathbf{Z}, t)\}$ and find the optimal $\mathbf{H}$, $\mathbf{H}_{\text{opt}}$, to minimize $\text{var}\{\widehat{\Delta}_{\text{LM}}(t) - n^{-1}\sum_{i=1}^{n}\{I(\mathcal{G}_i = B) - p\}\mathbf{H}(\mathbf{Z}_i, t)\}$. To approximate $\mathbf{H}_{\text{opt}}$, we first note that $\widehat{\mathcal{D}}_{\text{LM}}(t) = n^{\frac{1}{2}}\{\widehat{\Delta}_{\text{LM}}(t) - \Delta(t)\}$ is asymptotically equivalent to

$$\begin{aligned}&n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[\{\phi_{1i,A}(t,t_0)+\phi_{2i,A}(t,t_0)\}\frac{I(\mathcal{G}_i=A)}{(1-p)} - \{\phi_{1i,B}(t,t_0)+\phi_{2i,B}(t,t_0)\}\frac{I(\mathcal{G}_i=B)}{p}\right]\\ =&n^{-\frac{1}{2}}\sum_{i=1}^{n}\left[\phi_{1i,A}(t,t_0)+\phi_{2i,A}(t,t_0)-\phi_{1i,B}(t,t_0)-\phi_{2i,B}(t,t_0)-g_i(t,t_0)\{I(\mathcal{G}_i=B)-p\}\right]\end{aligned}$$

where

$$g_i(t,t_0) = \frac{\phi_{1i,A}(t,t_0)+\phi_{2i,A}(t,t_0)}{1-p} + \frac{\phi_{1i,B}(t,t_0)+\phi_{2i,B}(t,t_0)}{p}$$

It is straightforward to see that

$$\mathbf{H}_{\text{opt}}(\mathbf{Z}_i,t) = E\{g_i(t,t_0) \mid \mathbf{Z}_i\} = \sum_{\mathcal{G}\in\{A,B\}}(1-p)^{-I(\mathcal{G}=A)}p^{-I(\mathcal{G}=B)}E\{\phi_{1i}(t,t_0)+\phi_{2i}(t,t_0) \mid \mathbf{Z}_i, \mathcal{G}_i = \mathcal{G}\}.$$

To derive an expression for $\mathbf{H}_{\text{opt}}$, we note that

$$E\left\{I(X_{\mathbb{L}i} > t_0)S_{U_i}(t \mid t_0)\int_{t_0}^{t}\frac{dM_{C_i}(s|t_0)}{\pi_{U_i}(s|t_0)} \mid \mathbf{Z}_i\right\} = 0 = E\left\{S_{U_i^*}(t_0)\int_0^{t_0}\frac{dM_{C_i}(s)}{\pi_{U_i^*}(s)} \mid \mathbf{Z}_i\right\}$$

since $E\{M_{Ci}(s) \mid \mathbf{Z}_i, \mathcal{G}_i\} = 0 = E\{M_{C_i}(s|t_0) \mid U_i, X_{\mathbb{L}i} > t_0\} = 0$. Therefore,

$$\begin{aligned}
E\left\{\phi_{1i}(t,t_0) \mid \mathbf{Z}_i, \mathcal{G}_i\right\} &= E\left[\frac{I(X_{\mathbb{L}i} > t_0)}{S_{C,\mathcal{G}_i}(t_0)}\left\{\frac{I(X_{\mathbb{L}i} > t)}{S_{C,\mathcal{G}_i}(t|t_0)} - S_{\mathcal{G}_i}(t|t_0)\right\} \mid \mathbf{Z}_i, \mathcal{G}_i\right] \\
&= \{S_{\mathbf{Z}_i,\mathcal{G}_i}(t \mid t_0) - S_{\mathcal{G}_i}(t|t_0)\}S_{\mathbf{Z}_i,\mathcal{G}_i}(t_0) \\
E\left\{\phi_{2i}(t,t_0) \mid \mathbf{Z}_i, \mathcal{G}_i\right\} &= E\left\{S_{\mathcal{G}}(t|t_0)\left[\frac{I(X_{\mathbb{L}i} > t_0)}{S_{C,\mathcal{G}_i}(t_0)} - S_{\mathcal{G}_i}(t_0)\right] \mid \mathbf{Z}_i, \mathcal{G}_i\right\} \\
&= S_{\mathcal{G}_i}(t|t_0)\left\{S_{\mathbf{Z}_i,\mathcal{G}_i}(t_0) - S_{\mathcal{G}}(t_0)\right\}
\end{aligned}$$

where $S_{C,\mathcal{G}}(t) = P(C > t \mid \mathcal{G})$, $S_{\mathbf{Z},\mathcal{G}}(t,t_0) = P(T_{\mathbb{L}} > t \mid T_{\mathbb{L}} > t_0, \mathbf{Z}, \mathcal{G})$ and $S_{\mathbf{Z},\mathcal{G}}(t_0) = P(T_{\mathbb{L}} > t_0 \mid \mathbf{Z}, \mathcal{G})$. Therefore, $\mathbf{H}_{\text{opt}}$ can be chosen as

$$\mathbf{H}_{\text{opt}}(\mathbf{Z}, t) = -\sum_{\mathcal{G}\in A,B} \frac{S_{\mathbf{Z},\mathcal{G}}(t) - S_{\mathcal{G}}(t)}{(1-p)^{I(\mathcal{G}=A)}p^{I(\mathcal{G}=B)}} \tag{B·1}$$

Thus, this is the optimal basis for augmenting the proposed estimator. Note that $S_{\mathbf{Z},A}(t)$ for patients with $\mathcal{G}_i = B$ is the survival probability for his/her counterfactual survival time in treatment A.

Supplementary Appendix C: Interval Estimation via Resampling

Though we have derived the form of the variance of our proposed estimator in Appendix II, estimating this variance explicitly is computationally involved. To overcome this issue, we construct confidence intervals(CIs) using a perturbation-resampling method (Park & Wei, 2003; Cai et al., 2005; Tian et al., 2007) to approximate the distributions of the aforementioned estimators. Specifically, let $\{\mathbf{V}^{(b)} = (V_1^{(b)}, ..., V_n^{(b)})^{\intercal}, b = 1, ...B\}$ be $n \times B$ independent copies of a positive random variable V from a known distribution with unit mean and unit variance. Let $\widehat{\boldsymbol{\beta}}^{(b)}$ be the solution to

$$\sum_{i\in\Omega_{t_0}} \int_{t_0}^{\infty} V_i^{(b)} \left\{\mathbf{W}_i - \frac{\sum_{j\in\Omega_{t_0}} V_j^{(b)} Y_j(z) e^{\beta^{\intercal}\mathbf{W}_j}\mathbf{W}_j}{\sum_{j\in\Omega_{t_0}} V_j^{(b)} Y_j(z) e^{\beta^{\intercal}\mathbf{W}_j}}\right\} dN_i(z) = 0.$$

Let $\widehat{S}^{(b)}(t \mid t_0) = \frac{\sum_{i\in\Omega_{t_0}} V_i^{(b)} \exp\{-\widehat{\Lambda}^{(b)}_{\widehat{U}_i}(t)\}}{\sum_{i\in\Omega_{t_0}} V_i^{(b)}}$, $\widehat{\Lambda}^{(b)}_{\widehat{U}_i}(t) = \int_{t_0}^{t} \frac{\sum_{i\in\Omega_{t_0}} V_i^{(b)} K_h(\hat{D}^{(b)}_{ui}) dN_i(z)}{\sum_{i\in\Omega_{t_0}} V_i^{(b)} K_h(\hat{D}_{ui}) Y_i(z)}$, $\hat{D}^{(b)}_{ui} = \widehat{U}^{(b)}_i - u$ and $\widehat{U}^{(b)}_i = \widehat{\boldsymbol{\beta}}^{(b)\mathsf{T}}\mathbf{W}_i$. We may obtain $\widehat{S}^{(b)}(t_0)$ by replacing $\mathbf{W}_i = \mathbf{Z}_i$ throughout and using all patients and let $\widehat{S}^{(b)}_{\text{LM}}(t) \equiv \widehat{S}^{(b)}(t \mid t_0)\widehat{S}^{(b)}(t_0)$. Then one can estimate the variance of $\widehat{S}_{\text{LM}}(t)$ as the empirical variance of $\{\widehat{S}^{(b)}_{\text{LM}}(t), b = 1, ..., B\}$. This procedure can be used to obtain $\widehat{S}^{(b)}_{\text{LM},A}(t), \widehat{S}^{(b)}_{\text{LM},B}(t)$, and $\widehat{\Delta}^{(b)}_{\text{LM}}(t) = \widehat{S}^{(b)}_{\text{LM},A}(t) - \widehat{S}^{(b)}_{\text{LM},B}(t)$ for $b = 1, ...B$. Then one can estimate $\hat{\sigma}(\widehat{\Delta}_{\text{LM}}(t))$ as the empirical variance of $\widehat{\Delta}^{(b)}_{\text{LM}}(t)$. Let $\widehat{a} = \sum_i \{I(\mathcal{G}_i = B) - p\}\widehat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)$. To examine whether our proposed procedure benefits from augmentation, we examine $\widehat{\Delta}_{\text{AUG}}(t)$ by obtaining $\hat{\epsilon}$ as $\widehat{\text{var}}(\widehat{a})^{-1}\widehat{\text{cov}}\{\widehat{\Delta}_{\text{LM}}(t), \widehat{a}\}$ where $\widehat{\text{var}}(\widehat{a})$ is the empirical variance of $\widehat{a}^{(b)}$ and $\widehat{\text{cov}}\{\widehat{\Delta}_{\text{LM}}(t), \widehat{a}\}$ is the empirical covariance of $\widehat{\Delta}^{(b)}_{\text{LM}}(t)$ and $\widehat{a}^{(b)}$ for $b = 1, ...B$ and $\widehat{a}^{(b)} = \sum_i V_i^{(b)}\{I(\mathcal{G}_i = B) - p\}\widehat{\mathbf{H}}_{\text{opt}}(\mathbf{Z}_i, t)$. To construct CIs, one can either use the empirical quantiles of the perturbed samples or a normal approximation. Similar procedures can be used to obtain variance and interval estimates for $\widehat{S}_{\text{KM}}(t)$. The validity of the perturbation-resampling procedure can be shown using similar arguments as in Cai et al. (2010) and Zhao et al. (2010) since the distribution of $\sqrt{n}\{\widehat{S}_{\text{LM}}(t) - S(t)\}$ can be approximated by the distribution of $\sqrt{n}\{\widehat{S}^{(b)}_{\text{LM}}(t) - \widehat{S}_{\text{LM}}(t)\}$ conditional on the observed data.

References

Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics* pages 1071–1095.

Cai, T., Tian, L., Uno, H., Solomon, S. & Wei, L. (2010). Calibrating parametric subject-specific risk estimation. *Biometrika* **97**, 389–404.

Cai, T., Tian, L. & Wei, L. J. (2005). Semiparametric Box-Cox power transformation models for censored survival observations. *Biometrika* **92**, 619–632.

Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via kmt constructions. *The Annals of Probability* **17**, 266–291.

Park, Y. & Wei, L. J. (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika* **90**, 717–23.

Tian, L., Cai, T., Goetghebeur, E. & Wei, L. (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* **94**, 297–311.

Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Periodica Mathematica Hungarica* **8**, 53–55.

Zhao, L., Cai, T., Tian, L., Uno, H., Solomon, S. & Wei, L. (2010). Stratifying subjects for treatment selection with censored event time data from a comparative study. *Harvard University Biostatistics Working Paper Series* page 122.